INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY AND CREATIVE ENGINEERING (ISSN:2045-8711)
VOL.9 NO 8 AUGUST 2019

Effective Feature Selection Method for Cervical Cancer Dataset Using Data Mining Classification Analytical Model

Dr.D.Rajakumari[1], S.Karthika[2]
Assistant Professor, Department of Computer Science, Nandha Arts and Science College, Erode, Tamilnadu, India[1]
rsrajakumarid@gmail.com
MPhil Part-Time Scholar, Department of Computer Science, Nandha Arts and Science College, Erode, Tamilnadu, India[2]
Karthika109suresh@gmail.com

***Abstract*— This paper Effective Prediction Model for Cervical Cancer disease Using Data Mining Classification Algorithm describes classification techniques and shows the advantage of feature selection approaches to the best predicting of cervical cancer disease. There are 32 attributes with 858 samples. Besides, this data suffers of missing values and imbalance data. Therefore, over-sampling, under-sampling and imbedded over and under sampling have been used. In this paper implemented a feature model construction and comparative analysis for improving prediction accuracy of cervical cancer patients in four phases. In first phase, min-max normalization algorithm is applied on the original cervical cancer patient datasets collected from UCI repository. In cervical cancer dataset prediction second phase, by the use of feature selection, subset (data) of cervical cancer patient dataset from whole normalized cervical cancer patient datasets is obtained which comprises only significant attributes. Third phase, classification algorithms are applied on the data set. In the fourth phase, the accuracy will be calculated using root mean square value, root mean error value. KNN and SVM algorithm is considered as the better performance algorithm after applying feature selection. Finally, the evaluation is done based on accuracy values. Thus outputs shows from proposed GA base feature extraction with classification model implementations indicate that KNN and SVM algorithm performances all other classification algorithm with the help of feature selection with an accuracy of 97.60%.**

Keywords: Cervical Cancer dataset, Data Mining Algorithm, KNN, SVM

I. INTRODUCTION

Data Mining is one of the most encouraging areas of research with the purpose of finding useful information from voluminous data sets. It has been used in many domains like image mining, opinion mining, web mining, text mining, graph mining etc. Its applications include anomaly detection, financial data analysis, medical data analysis, social network analysis, market analysis etc,

Data Mining is particularly useful in medical field when no availability of evidence favouring a particular treatment option is found. Large amount of complex data is being generated by healthcare industry about patients, diseases, hospitals, medical equipment, claims, treatment cost etc. that requires processing and analysis for knowledge extraction. Data mining comes up with a set of tools and techniques which when applied to this processed data, provides knowledge to healthcare professionals for making appropriate decisions and enhancing the performance of patient management tasks.

Millions of early deaths among women is due to lung and breast cancer but cervical cancer is most treacherous because it is only diagnosed in females. Woman's reproductive system consists of cervix, uterus, vagina and the ovaries. Cervix is the opening to the uterus from the vagina where cervical cancer occurs [4]. Sexually transmitted human papillomavirus (HPV) is the important cause of cervical cancer.

Cervical Cancer occurrence is plentiful in low- and middle-income countries. The important task of cervical cancer is screening. A perfect screening test is the one that is least incursive, easy to accomplish, acceptable to subject, inexpensive and effective in diagnosing the

disease process in its early incursive stage when the treatment is easy for illness. There are four screening methods including cervical cytology also called Pap smear test, biopsy, Schiller and Hinslemann.

Cytology screening method is a microscopic analysis of cells scratched from the cervix and is used to detect cancerous or pre-cancerous conditions of the cervix. Biopsy method is a surgical process which includes finding of a living tissue sample for performing diagnosis. The solution of iodine has applied for visual inspection of cervix known as Hinslemann test. Lugol's iodine is used for visual assessment of cervix after smearing Lugol's iodine recognition rate of doubtful region over the cervix, this is also known as Schiller test.

This research work focuses on a prediction of disease. Since there are many related diseases, the Cervical Cancer disease is very dangerous because it leads to failure and also it cannot be predicted at early stages. The Cervical Cancer disease has stages which can be identified by regular checkup. If the disease is diagnosed than the patient's past history is analyzed. The classification model plays a vital role in the prediction of diseases. The aim of this research work is to develop an efficient predictive healthcare decision support system using data mining techniques. A common or dataset is trained in this system using KNN, MLP, SVM and Naïve Bayes classification algorithms and tested with the sample data which predict the patent's outcome of Cervical Cancer Diseases.

Data mining has been with success utilized in data discovery for prognostic functions to form a lot of active and correct call. Different data mining techniques i.e. Decision Tree, Bayesian Network, K-Nearest Neighbor, Naïve Bayes, Support Vector Machine, Multi layer perceptron etc. are used to predict disease in early stage which also helps to avoid the patient's complications. The main objective of this research work is to predict disease using Step wise Regression Model (SRM) and Built around the Random Forest Classification algorithm (BRFC), the result is obtained by comparing the algorithms and analysis the performance of the algorithm. Different data mining techniques are used to pull data. The experimental comparison of KNN, MLP, SVM and NBC are done based on the performance measures of classification accuracy and execution time.

II. RELATED WORKS

A. Ashfaq Ahmed et al., [1] have given a piece exploitation machine learning techniques, particularly Support Vector Machine [SVM] and Random Forest [RF]. These were wont to study, classify and compare cancer, liver and cardiovascular disease knowledge sets with variable kernels and kernel parameters. Results of Random Forest and Support Vector Machines were compared for various knowledge sets like carcinoma unwellness dataset, disease dataset and cardiovascular disease dataset. It's over that variable results were determined with SVM classification technique with completely different kernel functions.

B. Giovanni Caocci et al., [2] so as to predict future urinary organ Transplantation Outcome, they taken discrimination between a man-made Neural Network and supplying Regression. Comparison has been done supported the Sensitivity and specificity of supplying Regression and a man-made Neural Network within the prediction of urinary organ rejection in 10 coaching and corroborative datasets of urinary organ transplant recipients. From the experimental results that each the formula approaches were complementary and their combined algorithms won't to improve the clinical decision-making method and prognosis of urinary organ transplantation.

C. Lakshmi.K.R et al., [3] analyzed Artificial Neural Networks, call tree and Logical Regression supervised machine learning algorithms. These algorithms are used for urinary organ chemical analysis. For classification method they used an information mining tool named Tanagra. The tenfold cross validation is employed so as to gauge the classified knowledge proceeded by the comparison of these knowledge. From the experimental result they absorbed that ANN performed higher than the choice tree and Logical Regression algorithms.

D. Neha Sharma et al., [4] detected and expected urinary organ diseases as a prelude to correct treatment to patients. The system was used for detection in patients with disease and also the results of their IF-THEN rules expected the presence

of a disease. Their technique used fuzzy systems and a neural network referred to as a neural blur system, supported the results of the input file set obtained. Their system was a mix of fuzzy systems that created results exploitation correct mathematical calculations, rather than probabilistic based mostly classifications. Usually results supported arithmetic tends to possess higher accuracies. Their work was ready to acquire helpful knowledge in conjunction with optimizations in results.

E. Swathi Baby P et al., [5] n contestable that data processing strategies may be effectively employed in medical applications. Their study collected knowledge from patients affected with excretory organ diseases. The results showed knowledge mining's pertinence during a sort of medical applications. K-means (KM) rule will verify range of clusters in massive knowledge sets. Their study analyzed tree AD, J48, star K, theorem wise, random forest and tree - based ADT naive theorem on J48 renal disorder knowledge Se and noted that the techniques offer applied mathematics analysis on the utilization of algorithms to predict excretory organ diseases in patients.

G. Talha Mahboob Alam *et al.,* [6] in their study data mining techniques including decision tree algorithms are used in biomedical research for predictive analysis. Cervical cancer prediction through different screening methods using data mining techniques like Boosted decision tree, decision forest and decision jungle algorithms as well performance evaluation has done on the basis of Area under Receiver operating characteristic(AUROC) curve, accuracy, specificity and sensitivity.

H.Veenita Kunwar et al., [7] in their study had foreseen Cervical Cancer disorder (CKD) mistreatment naive theorem classification and artificial neural network (ANN). Their results showed that naive theorem created correct results than artificial neural networks. it had been conjointly ascertained that classification algorithms were wide used for investigation and identification of CKDs.

I. Vijayarani et al., [8] classification method is employed to classify four varieties of excretory organ diseases. Comparisons of Support Vector Machine (SVM) and Naïve mathematician classification algorithms are done supported the performance factors, classification, accuracy and execution time. As results, the SVM achieves enhanced classification performance. Therefore it's thought-about because the best classifier when put next with Naïve mathematician classifier rule. However, Naïve mathematician classifier classifies the information with minimum execution time. during this study, we tend to apply data processing techniques, recently hierarchic among the highest ten as best classifiers, to predict Cervical disorder on the idea of the data attributes within the info employed in order to reason patients World Health Organization are littered with the Cervical Cancer renal disorder (ckd) and patients World Health Organization don't seem to be littered with it (not ckd).

J. Dhayanand et al., [9] have conferred a piece to predict renal disorder by classifying four varieties of excretory organ diseases: Acute Nephritic Syndrome, Cervical Cancer renal disorder, acute failure and Cervical Cancer nephritis, mistreatment Support Vector Machine (SVM) and Artificial Neural Network (ANN), then examination the performance of these two algorithms on the idea of accuracy and execution time. The results show that the performance of the ANN is healthier than the SVM rule.

K. Sharma et al., [10] applied varied machine learning algorithms to a tangle within the domain of diagnosis and analysed their potency in predicting the results. The matter selected for the study is that the designation of the Cervical Cancer nephropathy. The dataset used for the study consists of four hundred instances and twenty four attributes. The authors evaluated twelve classifications on techniques by applying them to the Cervical Cancer nephropathy knowledge. So as to calculate potency, results of the prediction by candidate ways were compared with the particular medical results of the topic.

III. RESEARCH METHODOLOGY

In the paper system, a classical approach is papered for locating the diseases of urinary organ cancer victimization data processing classification techniques of Random Forest and Naïve mathematician. The techniques offer profit to the doctors, physicians, medical students and patients to form call relating to the diagnosing of the urinary organ cancer diseases.

The papered KNN primarily based classifier determines neighborhoods directly from coaching observations and it works with numeric feature vector of the urinary organ cancer dataset. The foremost advantage of this approach is that the correct operative coming up with supported image diagnostic information of the urinary organ cancer patients. The papered approach is employed to discover the urinary organ cancer patients affected and also the experimental application shows the results of the potency of the papered approach.

In addition to it for analyzing aid information, major steps information mining approaches like preprocess data, replace missing values, feature choice, machine learning and build call square measure applied on train dataset. Finally the random forest methodology has been dead on the coaching dataset of urinary organ cancer sickness for the classification method.

- Decision tree predicts a category victimization predefined classification tree with contains each numerical and categorical feature vector.
- To guarantee the validity of result's allotted by distribution varied values of K.
- The application is often utilized by anybody particularly for medical practitioners via web for diagnosing purpose.
- Select the class-outliers, that is, coaching information that square measure classified incorrectly by Random forest (for a given N time k th*)*

A. System Architecture

- Step1: Read the Cervical Cancer Dataset from UCI Machine learning Repository. The dataset have 400 records.
- Step 2: Normalize the Cervical patient dataset using Z-Score Normalization.
- Step 3: Feature extraction will be done by using Step wise Regression Model (SRM)

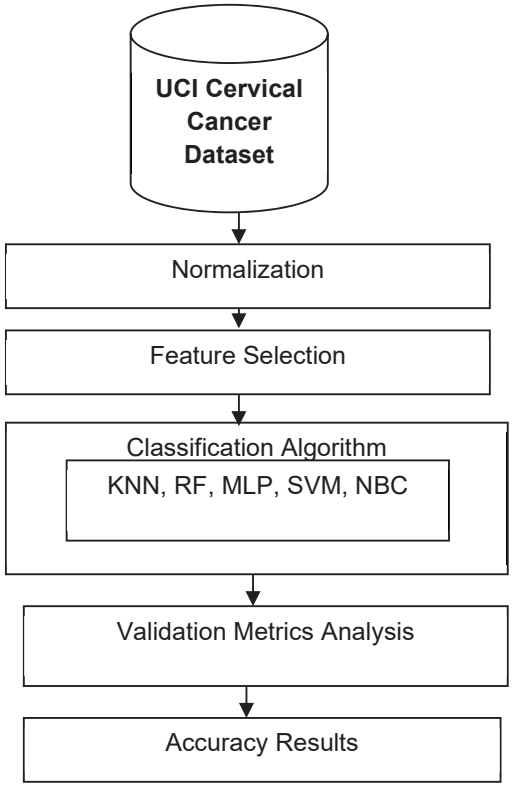

Fig 3.1 System Architecture

- Step4: The feature will be selected and put in to data frame.
- Step5: Classification algorithms are applied on the selected feature.
- Step6: KNN Classification to create centered point of data a new group contains the most important data points and others will be considered as outliers
- Step 7: RF classification, multiple trees are induced in the forest, the number of trees is pre-decided by the parameter N-tree.
- Step 8: SVM Classification to create a new group contains the most important data points and others will be considered as outliers.
- Step 9: NBC Classification prediction values for RPCC and RCC and compare train dataset and accuracy calculate.
- Step 10:To apply Cancer dataset using MLP classification model and accuracy calculated.

- Step 11: The results are obtained, MLP and SVM gives better accuracy when compare to other algorithms.
- Step 12: Accuracy will be analyzed.
- Step13: Finally the evaluation metrics will be calculated.

B. Normalization

Normalization is scaling technique or a pre process stage. Where, we are able to discover new dimension from associate degree existing one series. It is often useful for the prediction or statement operates heaps. Therefore maintain the big distinction of prediction and statement the standardization technique is needed to form them nearer.

Z-score is that the variety of normal deviations from the mean an information purposes. However additional technically it's calculated of what percentage normal deviations below or on top of the population means that a rough score. A z-score is additionally referred to as a customary score and it are often placed on a standard distribution curve. Z-scores vary from -3 normal deviations (which would fall to the left of the conventional distribution curve) up to +3 normal deviations (which would fall to the way right of the conventional distribution curve). so as to use a z-score, you wish to spot the mean μ and conjointly the population variance σ.

$$z = (x – \mu) / \sigma$$

C. Feature Selection

Feature extraction is that the model of choosing a set of the terms gift within the coaching set and victimization solely this set as options in text classification. Feature extractions provide 2 main functions. First, it makes coaching and applying a classifier additional powerful by decreasing the scale of the adequate vocabulary. Feature extraction method is of explicit significance for classifiers that, unlike NB, square measure costly to coach. Second, feature extraction typically will increase classification accuracy by eliminating noise options. A noise feature is one that, once joined to the document illustration, will increase the classification error on new knowledge. Facilitating knowledge visual image is dashing up the execution of mining algorithms and reducing descending times

Stepwise regression n may be a combination of the forward and backward choice techniques. Stepwise regression may be a modification of the forward choice in order that when every step within which a variable was added , all candidate variables within the model square measure checked to examine if their significance has been reduced below the required tolerance level. If a no important variable is found, it's aloof from the model. Stepwise regression needs 2 significance levels: one for adding variables and one for removing variables. The cutoff likelihood for adding variables ought to be but the cutoff likelihood for removing variables in order that the procedure doesn't get into associate degree infinite loop.

D. Classification Algorithm

a) RF Algorithm

RF is an algorithmic program accustomed manufacture a choice tree that is increase of previous ID3 calculation. It en-large the ID3 algorithmic program is managing each continuous and distinct property, missing values and pruning trees once construction. The choice trees created by C4.5 are often used for grouping and sometimes cited as a applied math classifier. C4.5 creates call trees from a group of coaching urinary organ information same approach as Id3 algorithmic program. Because it could be a supervised learning algorithmic program it needs a group of coaching examples which may be seen as a pair: input object and a desired output worth (class). The algorithmic program analyzes the coaching set and frame a classifier that has to have the dimensions to accurately prepare each coaching and take a look at cases

b) NBC Model:

The Naive Bayesian classifier relies on Bayes' theorem with independence assumptions between predictors. Naive Thomas Bayes classifiers area unit a family of straightforward probabilistic classifiers supported applying theorem. Thomas Bayes theorem provides some way of conveying the posterior likelihood, P(c/x), from P(c), P(x), and P(x/c). It assumes that the result of the worth of a predictor (x) on a given class (c) is freelance of the values of alternative predictors. This assumption is named category conditional independence. The Naïve Bayesian classification predicts that the tuple 'x' belongs to the category 'c' victimization the formula.

$$P(c/x)= (x/c) / (P (x)$$

- P (c/x) is that the posterior likelihood of sophistication (target) given predictor (attribute).
- P(c) is that the previous likelihood of sophistication.
- P (x/c) is that the chance that is that the likelihood of predictor given category.
- P(x) is that the previous likelihood of predictor.

c) KNN Classification

K-Nearest Neighbor (Knn) –Techniques KNN could be a supervised learning algorithmic program that classifies new information supported minimum distance from the new information to the K nearest neighbor. The papered work has used geometrician Distance to outline the closeness. Pseudo-code for the KNN classifier is declared below:

- Step 1: Input: D= x=(x1,…..,xn) new instance to be classified • Step 2: for every labeled instance (xi,ci) Calculated(xi, x)
- Step 3: Ordered(xi, x) from lowest to highest, (i=1,….,N)
- Step 4: Select the K nearest instances to x : Dx K
- Step 5: Assign to x the foremost frequent category in Dx K

d) MLP (Multilayer Perceptron)

A multilayer perceptron (MLP) could be a feed forward artificial neural network model that maps urinary organ datasets of input file onto a collection of applicable outputs. Associative MLP classification could be a multiple layer of nodes in a much-directed graph, with every layer totally connected to following one. A side from the input nodes, every node could be a nerve cell (or process element) with a nonlinear activation perform. MLP classification urinary organ dataset utilizes a supervised learning technique known as back propagation for coaching urinary organ the network. MLP could be a modification of the quality linear perceptron and might distinguish knowledge that isn't linearly dissociable urinary organ dataset method.

E. ANALYSIS METRIC

Mean Absolute Error

Statistical exactness metrics valuate the accuracy of a system by examination the numerical recommendation scores against the particular user ratings for the user-item pairs within the take a look at dataset. Mean Absolute Error (MAE) between ratings and predictions could be a wide used metric

Root Mean sq. Error

The Root Mean sq. Error (RMSE) (also known as the foundation mean sq. deviation, RMSD) could be a oftentimes used live of the distinction between values expected by a model and therefore the values truly determined from the setting that's being modeled. These individual variations also are known as residuals, and therefore the RMSE serves to combination them into one live of prognostic power. The RMSE of a model prediction with relevancy the calculable variable X model is outlined because the root of the mean square error:

Root Relative Squared Error

Correlation – usually measured as a parametric statistic – indicates the strength and direction of a linear relationship between 2 variables (for example model output and determined values). Variety of various coefficients square measure used for various things

Kappa Metrics

It returns the constant value. It measures the agreement between classification and truth values. It of one represents good agreement, whereas a price of zero represents no agreement.

IV. EXPERIMENTAL RESULTS

A. Dataset description

Cervical cancer data involves 858 samples and 32 features as well as four classes (Hinselmann, Schiller, Cytology and Biopsy) has been published in. This paper focuses on studying the Biopsy target as it recommended by the literature review.

Attribute	Type	Attribute	Type	Attribute	Type
Age	Integer	STDs	Bool	STDs:HIV	Bool
Number of sexual partners	Integer	STDs (number)	Integer	STDs:Hepatitis B	Bool
First sexual intercourse (age)	Integer	STDs:condylomatosis	Bool	STDs:HPV	Bool
Number of pregnancies	Integer	STDs:cervical condylomatosis	Bool	STDs: Number of diagnosis	Integer
Smokes	Bool	STDs:vaginal condylomatosis	Bool	STDs: Time since first diagnosis	Integer
Smokes (years)	Bool	STDs:vulvo-perineal condylomatosis	Bool	STDs: Time since last diagnosis	Integer
Smokes (packs/year)	Bool	STDs:syphilis	Bool	Dx:Cancer	Bool
Hormonal Contraceptives	Bool	STDs:pelvic inflammatory disease	Bool	Dx:CIN	Bool
Hormonal Contraceptives (years)	Integer	STDs:genital herpes	Bool	Dx:HPV	Bool
IUD	Bool	STDs:molluscum contagiosum	Bool	Dx	Bool
IUD (years)	Integer	STDs:AIDS	Bool		

Fig 4.1. Cervical Cancer Dataset Attribute

B. Performance Analysis

Table 4.1 describes a training dataset for cervical cancer dataset classification for NBC Model and SVM analysis model. The table contains precision, recall, F-measure and accuracy details are shown

TABLE 4.1 CERVICAL TRAINING DATASET METRICS ANALYSIS

Techniques	SRC Feature		Precision	Recall	F-measure	Accuracy
	Instances	No of Attributes				
NBC	500	32-Including class Label	0.7152	0.7642	0.7772	0.7821
SVM (FSDMM)	500	12 - Including class Label	0.7554	0.8035	0.8182	0.8045

Fig 4.2 describes a training dataset for cervical cancer dataset classification for NBC Model and SVM analysis model. The figure contains precision, recall, F-measure and accuracy details are shown.

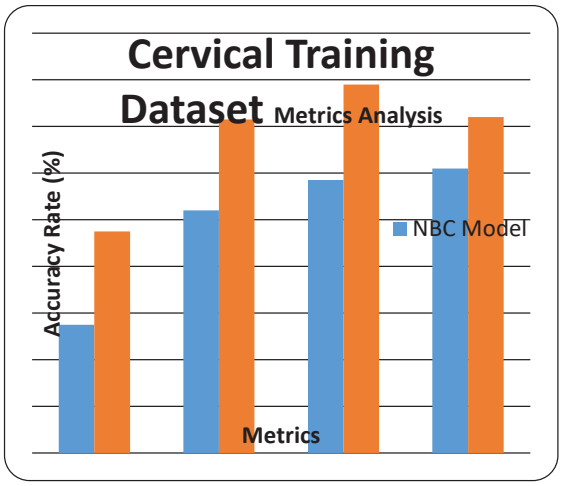

Fig 4.2 Cervical Training Dataset Metrics Analysis

Table 4.2 describes a test dataset for cervical cancer dataset classification for NBC Model and FSDMM analysis model. The table contains precision, recall, F-measure and accuracy details are shown.

TABLE 4.2 CERVICAL TEST DATASET METRICS ANALYSIS

Techniques	SRC Feature		Precision	Recall	F-measure	Accuracy
	Instances	No of Attributes				
NBC	250	32-Including class Label	0.7092	0.7565	0.7656	0.7981
SVM (FSDMM)	250	10 - Including class Label	0.7333	0.7964	0.8072	0.8323

Fig 4.3 describes a test dataset for cervical cancer dataset classification for NBC Model and FSDMM analysis model. The figure contains precision, recall, F-measure and accuracy details are shown.

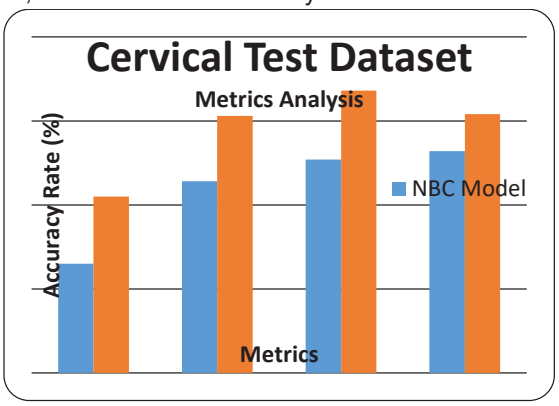

Fig 4.3 Cervical Test Dataset Metrics Analysis

In this table 4.3 describe time efficient analysis for Cervical cancer prediction model. In this table contain number of dataset, average time for execution for cancer prediction model details are shown,

Table 4.3 Time Analysis for NBC and FSDMM Model using Cervical Cancer Dataset

Number of Dataset	Number of Attribute	NBC Model (ms)	FSDMM Model (ms)
150	29	0.233	0.192
250	24	0.345	0.203
350	25	0.456	0.335
400	24	0.522	0.418
450	22	0.633	0.553
600	12	0.693	0.592

In this figure 4.3 describe time efficient analysis for Cervical cancer prediction model. In this figure contain number of dataset, average time for execution for cancer prediction model details are shown,

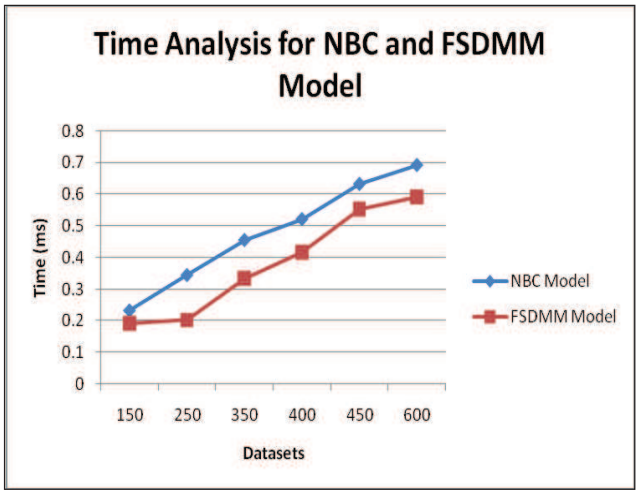

Fig 4.4Time Analysis for NBC and FSDMM Model using Cervical Cancer Dataset

In this Table 4.4 describe performance analysis for cervical cancer prediction model. In this figure contain number of dataset, average dataset prediction for cancer prediction model details are shown,

Table 4.3 Performance Analysis for NBC and FSDMM Model using Cervical Cancer Dataset

Number of Dataset	Number of Attribute	NBC Model (%)	FSDMM Model (%)
150	29	77.33	76.55
250	24	79.68	80.23
350	25	81.67	82.44
400	24	82.04	82.89
450	22	83.78	84.67
600	12	85.66	87.78

IN THIS FIGURE 4.5 DESCRIBE PERFORMANCE ANALYSIS FOR CERVICAL CANCER PREDICTION MODEL. IN THIS FIGURE CONTAIN NUMBER OF DATASET, AVERAGE DATASET PREDICTION FOR CANCER PREDICTION MODEL DETAILS ARE SHOWN

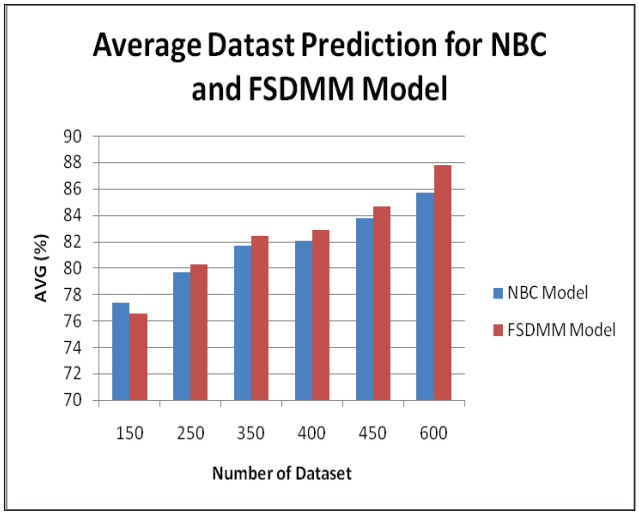

Fig 4.5 Performance Analysis for NBC and FSDMM Model using Cervical Cancer Dataset

V. CONCLUSION

In paper feature selection is done with the help of SRS approach. The whole datasets of cervical cancer patients is comprised of all relevant or irrelevant attributes. By the use of feature selection, a subset (data) of cervical cancer patient from whole cervical cancer patient datasets will be obtained which comprises only significant attributes

This result in the selection of 32 significant attributes consists of values of different classification algorithms. Comparison is made among classification algorithms out of which NBC and SVM algorithm is considered as the better performance algorithm. Because it gives higher accuracy in respective to other classification algorithms after applying feature selection: with an accuracy of 87.78%. The proposed methodology is used to predict the cervical cancer region into separable compartments. However, the method requires further improvement mostly regarding feature selection and segmentation of the cervical dataset into multiple components: renal cortex, renal column, renal medulla and renal pelvis. In addition this paper can be employed for detecting the heart diseases in future with the heart and liver dataset and classification of the diseases.

REFERENCES

[1] *Ashfaq Ahmed, K., Aljahdali, S., Hussain, S.N.: "Comparative prediction performance with support vector machine and random forest classification techniques", International Journal Computer Applications. 69 (11), 12–16 ,2016.*

[2] *Giovanni Caocci, Roberto Baccoli, Roberto Littera, Sandro Orrù, Carlo Carcassi and Giorgio La Nasa, "Comparison Between an Artificial Neural Network and Logistic Regression in Predicting Long Term Kidney Transplantation Outcome", Chapter 5, an open access article distributed under the terms of the Creative Commons Attribution License, http://dx.doi.org/10.5772/53104,2017.*

[3] *Lakshmi. K.R, Nagesh. Y and VeeraKrishna. M "Performance Comparison of Three Data Mining Techniques for Predicting Kidney Dialysis Survivability", International Journal of Advances in Engineering & Technology, Mar., Vol. 7, Issue 1, pg no. 242-254, 2016.*

[4] *Neha Sharma, Er. Rohit Kumar Verma, "Prediction of Kidney Disease by using Data Mining Techniques", Prediction of Kidney Disease by using Data Mining Techniques, 2016.*

[5] *Swathi Baby P and Panduranga Vital T, "Statistical Analysis and Predicting Kidney Diseases using Machine Learning Algorithms", International Journal of Engineering Research & Technology (IJERT), 2015.*

[6] *Talha Mahboob Alam,Muhammad milhan afzal khan,"Cervical Cancer Prediction through different screening methods using data mining",International Journal of Advanced Computer Science and Applications(IJACSA),2019.*

[7] *Veenita Kunwar, Khushboo Chandel, A. Sai Sabitha, and Abhay Bansal, "Chronic Kidney Disease Analysis Using Data Mining Classification Techniques", IEEE, 2016.*

[8] *Vijayarani, S., Dhayanand, S.: "Data mining classification algorithms for kidney disease prediction", International Journal of Cybern. Inf. (IJCI) 4(4), 13–25, 2017.*

[9] *Vijayarani, S., Dhayanand, S.: "Kidney disease prediction using SVM and ANN algorithms", International Journal Comput. Business Res. 6(2), 2017.*

Sharma, S., Sharma, V., & Sharma, A. (2017). Performance Based Evaluation of Various Machine Learning Classification Techniques for Chronic Kidney Disease Diagnosis. arXiv preprint arXiv:1606.09581,2017.